Comprendere il meccanismo di asserzione PIM

Sommario

Introduzione

Questo documento descrive il meccanismo di asserzione Protocol Independent Multicast (PIM), si concentra sui criteri di asserzione dei vincitori del PIM e approfondisce alcuni casi d'angolo.

Prerequisiti

Requisiti

Cisco raccomanda la conoscenza del meccanismo di asserzione PIM.

Componenti usati

Il riferimento delle informazioni contenute in questo documento è Cisco CSR1000V versione 16.4.1

Le informazioni discusse in questo documento fanno riferimento a dispositivi usati in uno specifico ambiente di emulazione. Su tutti i dispositivi menzionati nel documento la configurazione è stata ripristinata ai valori predefiniti. Se la rete è operativa, valutare attentamente eventuali conseguenze derivanti dall'uso dei comandi.

Che cos'è il meccanismo di asserzione PIM?

Quando in un segmento condiviso sono presenti più router abilitati per PIM, è possibile che questi router incontrino traffico multicast duplicato. Ciò si può verificare perché due o più router sullo stesso segmento condiviso possono avere una voce valida (S,G) o (*,G) che popola l'interfaccia in uscita verso il segmento condiviso per lo stesso gruppo IP/di destinazione di origine.

Il meccanismo di asserzione PIM viene utilizzato per rilevare ed eliminare la duplicazione del traffico multicast su un segmento condiviso. È importante notare che questo meccanismo non impedisce la duplicazione, ma utilizza la duplicazione del traffico multicast come trigger per

attivare questo meccanismo che seleziona un singolo server di inoltro per il flusso.

In caso di duplicazione del traffico multicast su un segmento condiviso, è possibile presupporre che più router inviino lo stesso messaggio (S,G) o (*,G) su un segmento condiviso. Se si sceglie un router per inoltrare il flusso in modo efficace, viene eliminata la duplicazione.

PIM sfrutta i messaggi di asserzione PIM che vengono attivati quando si riceve un pacchetto multicast nell'elenco interfacce in uscita (OIL). Questi messaggi di asserzione contengono metriche che vengono utilizzate per calcolare chi vincerà l'asserzione. I router downstream sulla LAN ricevono anche messaggi di asserzione PIM. Questi messaggi vengono quindi utilizzati dai dispositivi downstream per inviare i messaggi Join/Prune appropriati al router upstream che ha vinto la selezione dell'asserzione.

Scenario 1. Motivazione LHR

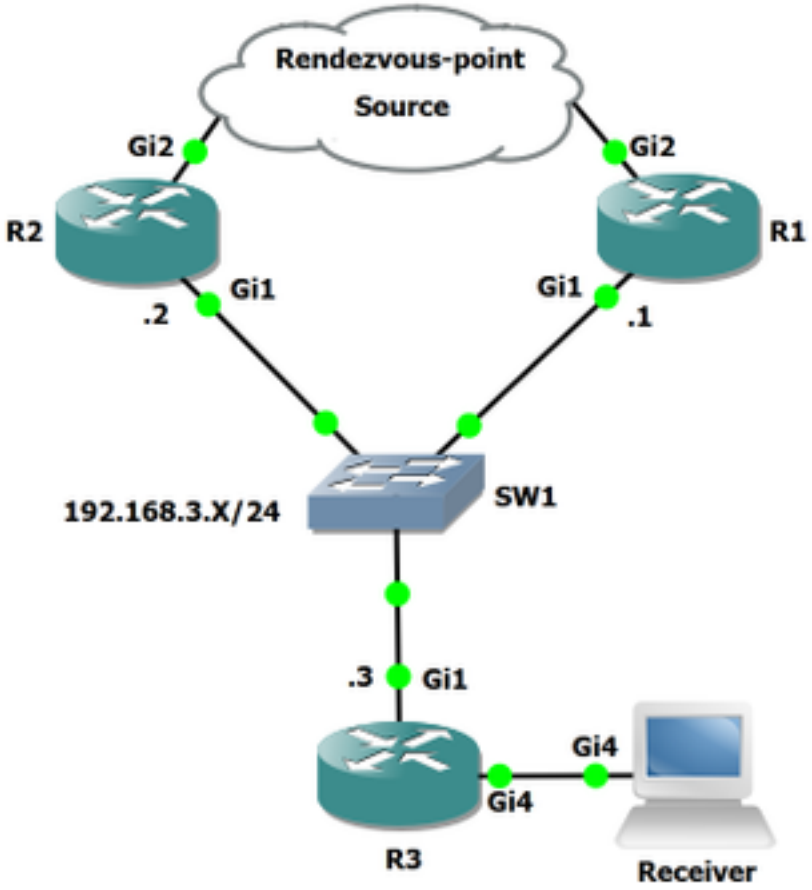


Figura 1.

Nel diagramma di rete, R3 è il router LHR (Last Hop Router), R3 si connette sia a R2 che a R1 tramite un segmento condiviso.

Quando si riceve un report IGMP (Internet Group Management Protocol) dal destinatario, R3 controlla chi è il router adiacente RPF verso l'RP. Nella topologia, R1 è l'RPF adiacente verso l'RP, quindi R3 invia un join (*,G) verso R1. Una volta che R1 si sposta verso il basso nel flusso (supponendo che il gruppo sia attivo) R3 invia un join (S,G) verso l'origine e abbassa l'albero di origine. R2 è l'RPF adiacente verso l'albero di origine, il che significa che R3 invierà il join (S,G) verso R2. R3 ha la stessa interfaccia RPF sia verso RP che verso l'origine. Qui è possibile vedere la tabella di route R3 per il gruppo 239.1.1.1.

```
R3#show ip mroute
IP Multicast Routing Table
Outgoing interface flags: H - Hardware switched, A - Assert winner, p - PIM Join
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode
(*, 239.1.1.1), 00:00:55/stopped, RP 192.168.0.100, flags: SJC
  Incoming interface: GigabitEthernet1, RPF nbr 192.168.3.1
  Outgoing interface list:
    GigabitEthernet4, Forward/Sparse, 00:00:55/00:02:04

(10.0.0.2, 239.1.1.1), 00:00:52/00:02:07, flags: JT
  Incoming interface: GigabitEthernet1, RPF nbr 192.168.3.2, Mroute
  Outgoing interface list:
    GigabitEthernet4, Forward/Sparse, 00:00:52/00:02:07

(*, 224.0.1.40), 00:01:22/00:02:09, RP 192.168.0.100, flags: SJPCL
  Incoming interface: GigabitEthernet1, RPF nbr 192.168.3.1
```

Come si può vedere in R3, il router RPF adiacente (*,G) è 192.168.3.1 e il router RPF adiacente verso il router (S,G) è 192.168.3.2. **Ciò dovrebbe far sì che sia R1 che R2 abbiano un OIL valido verso R3**. Di seguito sono riportate le voci:

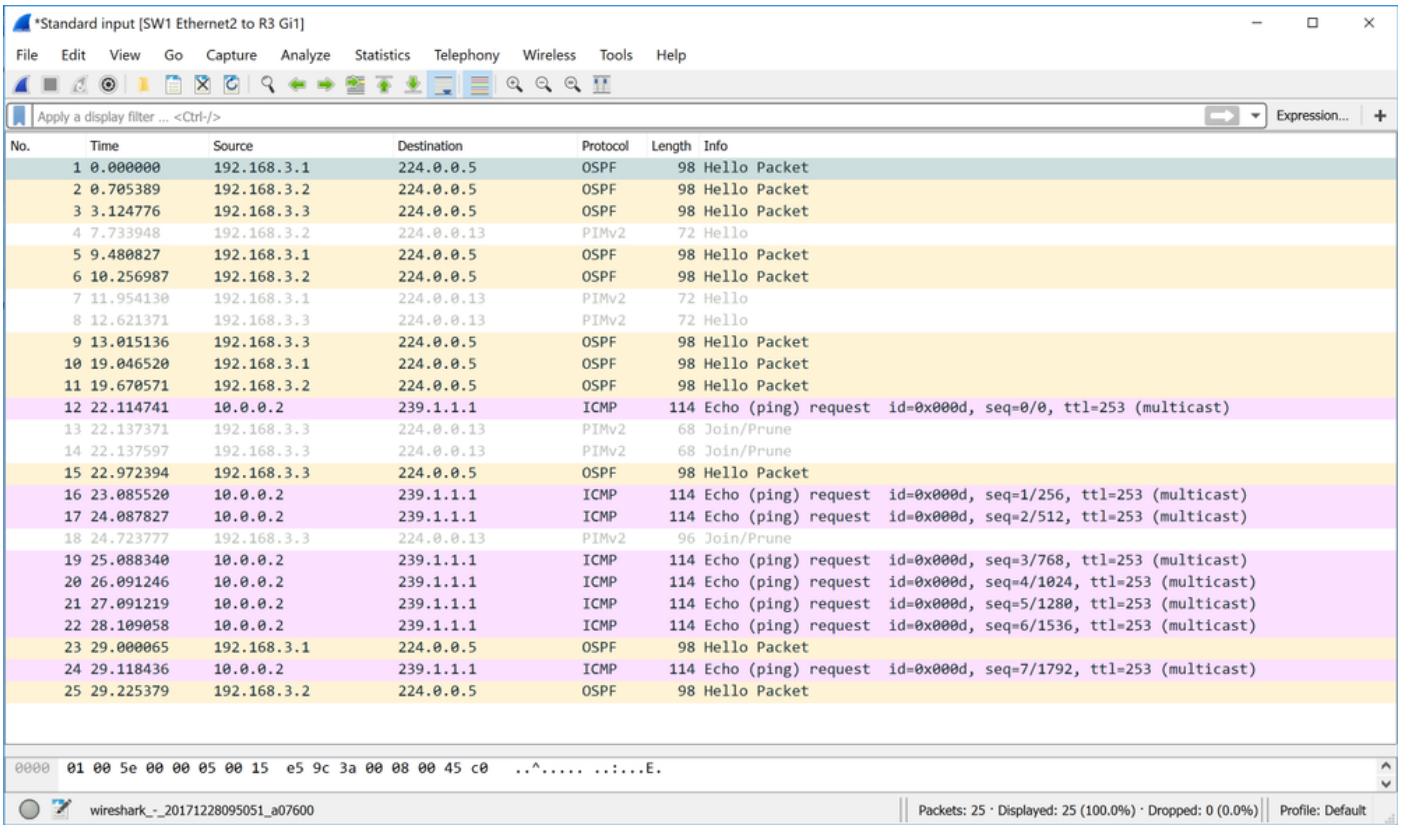
```
R1#show ip mroute
Outgoing interface flags: H - Hardware switched, A - Assert winner, p - PIM Join
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 239.1.1.1), 00:15:02/00:02:33, RP 192.168.0.100, flags: S
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.5.2
  Outgoing interface list:
    GigabitEthernet1, Forward/Sparse, 00:15:02/00:02:33

(10.0.0.2, 239.1.1.1), 00:13:24/00:02:33, flags: PR
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.5.2
  Outgoing interface list: Null

(*, 224.0.1.40), 00:29:17/00:02:51, RP 192.168.0.100, flags: SJCL
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.5.2
  Outgoing interface list:
    GigabitEthernet1, Forward/Sparse, 00:16:06/00:02:51
  Outgoing interface list: Null
```

```
R2#show ip mroute
IP Multicast Routing Table
Outgoing interface flags: H - Hardware switched, A - Assert winner, p - PIM Join
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 239.1.1.1), 00:08:00/stopped, RP 192.168.0.100, flags: SP
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.4.1
  Outgoing interface list: Null

(10.0.0.2, 239.1.1.1), 00:00:03/00:02:56, flags: T
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.4.1
  Outgoing interface list:
    GigabitEthernet1, Forward/Sparse, 00:00:03/00:03:26

(*, 224.0.1.40), 01:37:30/00:02:22, RP 192.168.0.100, flags: SJPL
```

```
Incoming interface: GigabitEthernet2, RPF nbr 192.168.4.1
```

Come accennato in precedenza, l'asserzione può essere attivata quando vi sono due router a monte con un valore OIL valido inserito in un segmento condiviso. Poiché sia R1 che R2 dispongono di un OLIO valido, controllare se è presente un meccanismo di asserzione nell'acquisizione dei pacchetti.

Questa acquisizione è stata acquisita sull'interfaccia R3 Gi1 verso SW1.



In questa acquisizione di pacchetti non vengono visualizzati pacchetti di asserzione anche se esistono tutti i prerequisiti per creare la duplicazione nel segmento condiviso tra R1, R2 e R3. Perché non vengono visualizzati pacchetti di asserzione PIM quando il flusso (S,G) è stato attivato?

Sembra che la RFC 7761 possa contenere la risposta a queste domande.

Riassunto della RFC 7761, sezione 4.2.2.

```
4.2.2.  Setting and Clearing the (S,G) SPTbit

   Basically, Update_SPTbit(S,G,iif) will set the SPTbit if we have the
   appropriate (S,G) join state, and if the packet arrived on the
   correct upstream interface for S, and if one or more of the following
   conditions apply:

   1.  The source is directly connected, in which case the switch to the
       SPT is a no-op.

   2.  The RPF interface to S is different from the RPF interface to the
       RP.  The packet arrived on RPF_interface(S), and so the SPT must
       have been completed.
```

```
3.  No one wants the packet on the RP tree.

4.  RPF'(S,G) == RPF'(*,G).  In this case, the router will never be
    able to tell if the SPT has been completed, so it should just
    switch immediately.  The RPF'(S,G) != NULL check ensures that the
    SPTbit is set only if the RPF neighbor towards S is valid.
```

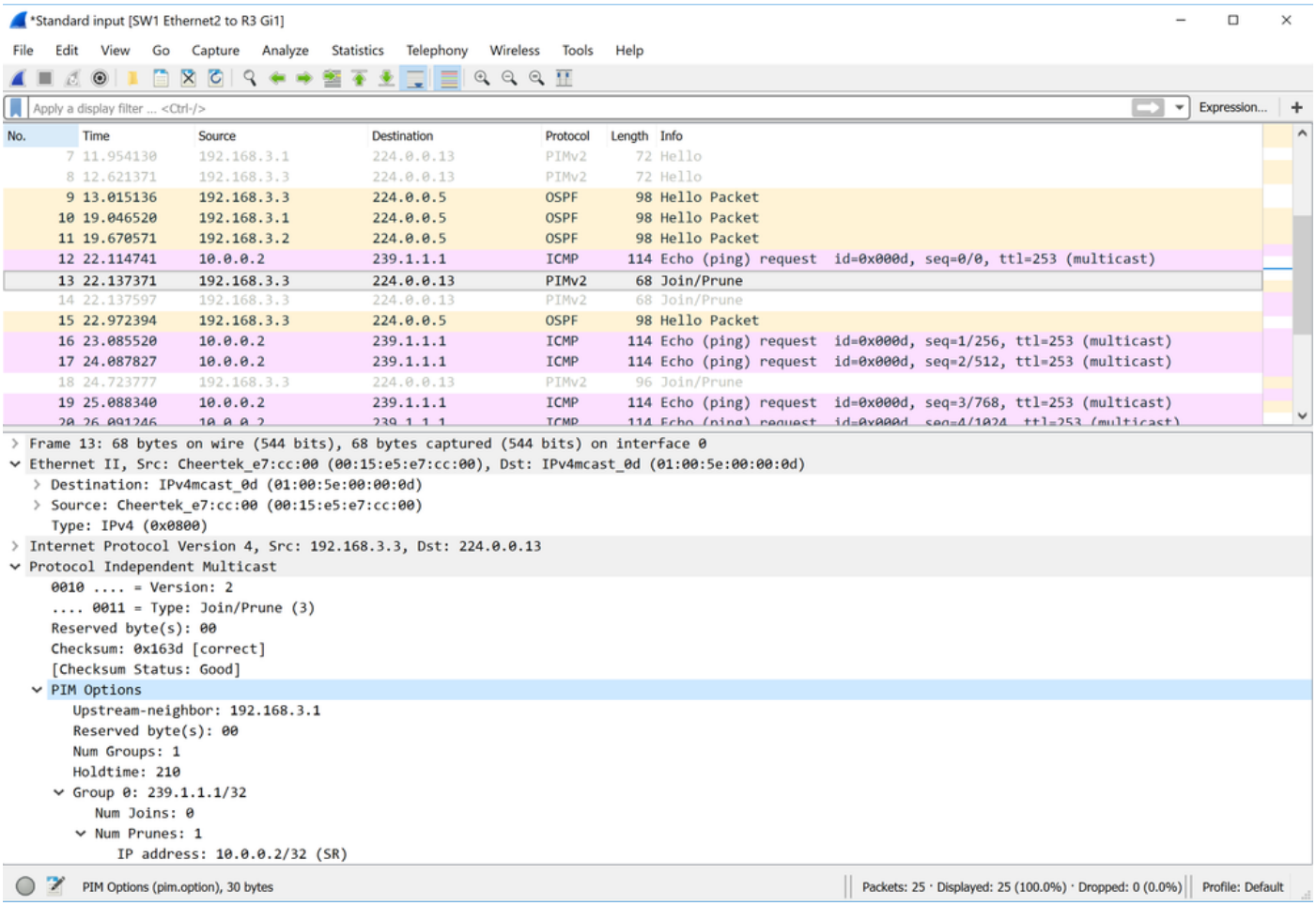
In the case where the RPF interface is the same for the RP and for S, but RPF'(S,G) and RPF'(*,G) differ, we wait for an Assert(S,G), which indicates that the upstream router with (S,G) state believes the SPT has been completed.

Il bit SPT (S,G) viene utilizzato per distinguere se inoltrare nello stato (*,G) o on (S,G). Quando si passa dall'albero RP all'albero di origine, si verifica un periodo di transizione in cui i dati arrivano a causa dello stato upstream (*,G) mentre lo stato upstream (S,G) è stabilito. In quel momento, il router deve continuare a inoltrare solo lo stato on (*,G). Ciò impedisce che si verifichino buchi neri temporanei causati dall'invio di una Prune(S,G,rpt) prima che lo stato a monte (S,G) sia stato definito.

Anche se sembra che lo scenario possa essere correlato all'ultimo punto menzionato in precedenza. Nel caso in cui l'interfaccia RPF è la stessa per l'RP e per S, ma RPF'(S,G) e RPF'(*,G) differiscono, attendiamo un'asserzione(S,G), che indica che il router a monte con stato (S,G) ritiene che l'SPT sia stato completato.

Affinché l'asserzione venga attivata, il router deve ricevere un pacchetto duplicato sul suo OIL già popolato per lo stesso gruppo IP/di destinazione di origine sul segmento. R3 è anche un LHR, ossia è designato per passare da (*,G) a SPT (S,G) quando un pacchetto viene ricevuto da (*,G).

Nell'acquisizione dei pacchetti si osserva che non vengono attivate asserzioni. Anche se vediamo una prugna inviata immediatamente dopo la ricezione della prima eco ICMP.

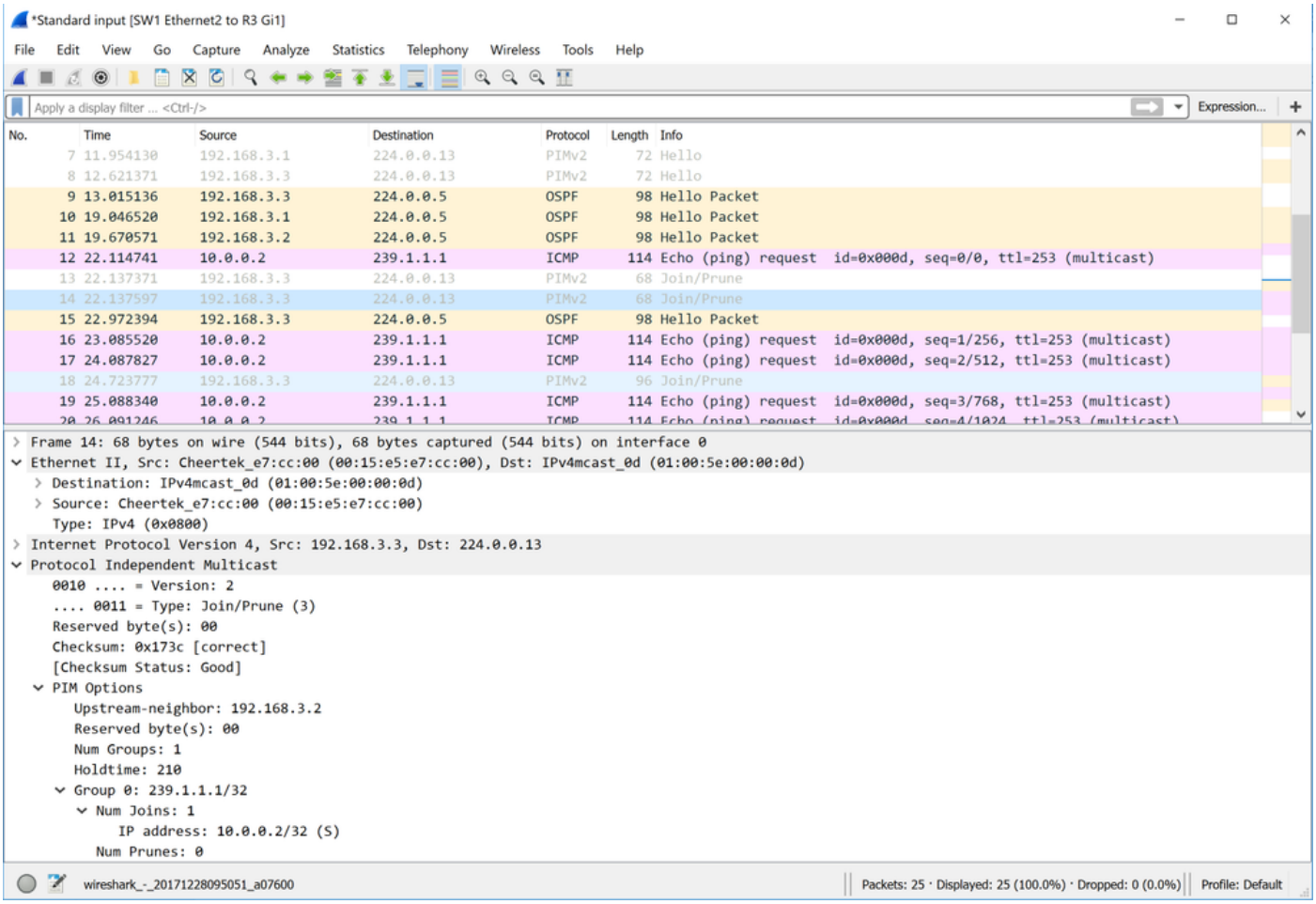


Come si può vedere, una volta ricevuto il primo pacchetto di richiesta ICMP (Internet Control Message Protocol) sull'interfaccia R3 G1, viene inviata una prugna (*,G) SR-bit verso la porta adiacente a monte 192.168.3.1. Questa prugna (*,G) viene eliminata per l'origine specifica definita.

Èpossibile visualizzare anche i seguenti flag impostati: (SR):

```
The S flag: indicates that the multicast group is a sparse mode group.
The R flag: The R flag is the RP-bit flag and indicates that the information in the (S, G) entry
is applicable to the shared tree.
```

Nel secondo pacchetto PIM n. 14, si nota che R3 cerca di unirsi alla struttura (S,G).



Una volta ricevuto il primo piano dati, il pacchetto R3 elimina (*,G) e crea (S,G). Questo è il motivo per cui non vengono visualizzati i pacchetti di asserzione PIM. Questo scenario è valido quando un LHR ha la stessa interfaccia RPF per (S,G) e (*,G). Sebbene questo comportamento possa essere leggermente diverso dalla RFC 7761, non deve causare problemi.

Proseguiamo ora con lo Scenario 2., il diagramma di questo scenario può essere visto qui:

Scenario 2. Selezione percorso asserzione

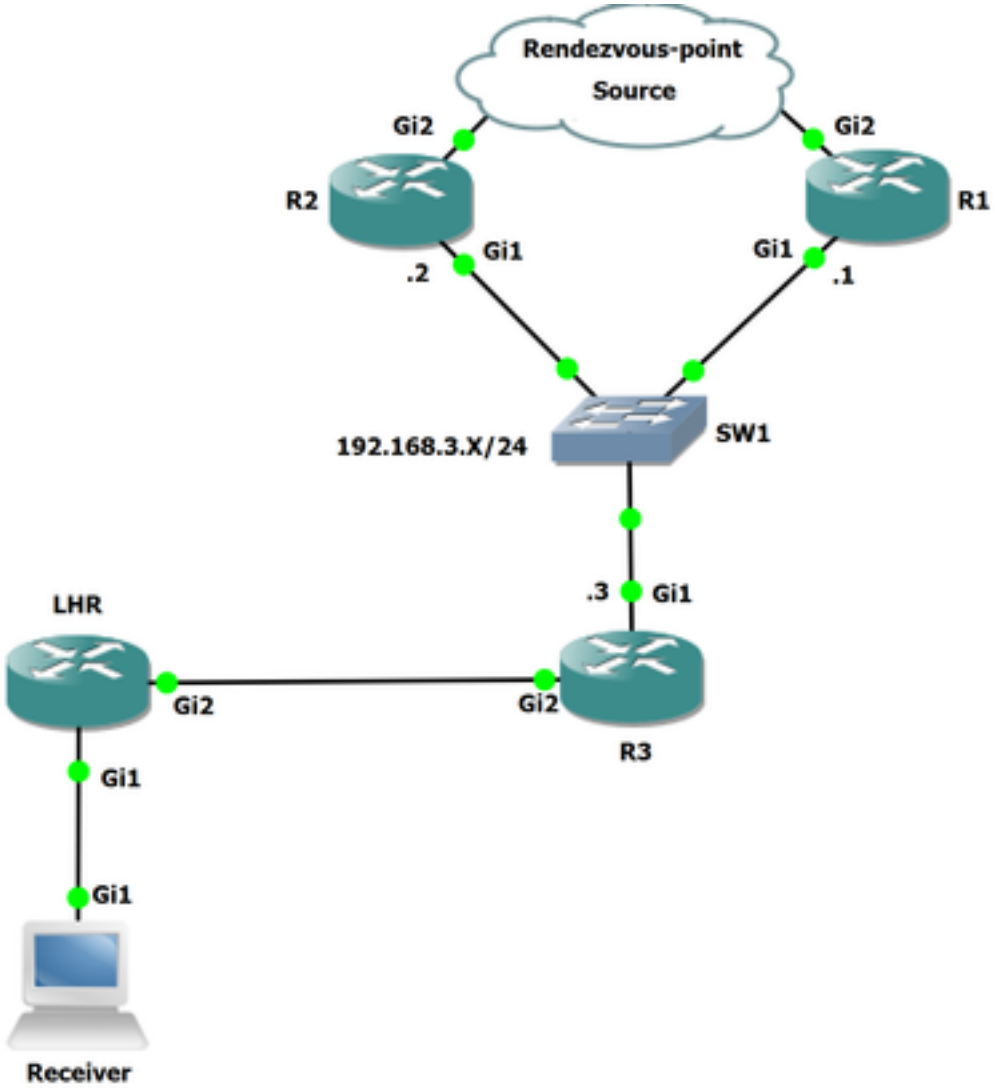


Figura 2.

In questa topologia, su R3 è connesso un altro router, ossia l'LHR. L'LHR si collega direttamente al ricevitore. L'origine e l'RP sono entrambe al di sopra di R2 e R1. L'RPF adiacente su R3 verso l'RP è R1 e l'RPF adiacente verso l'origine è R2.

Controllare il router adiacente RPF per l'origine e l'RP.

Qui è possibile vedere il vicino RPF in direzione dell'RP: 192.168.0.100 è 192.168.3.1.

```
R3#show ip rpf 192.168.0.100
RPF information for ? (192.168.0.100)
  RPF interface: GigabitEthernet1
  RPF neighbor: ? (192.168.3.1)
  RPF route/mask: 192.168.0.100/32
  RPF type: unicast (ospf 1)
  Doing distance-preferred lookups across tables
  RPF topology: ipv4 multicast base, originated from ipv4 unicast base
```

Qui si vede il vicino RPF verso la sorgente: 10.0.0.2 è 192.168.3.2.

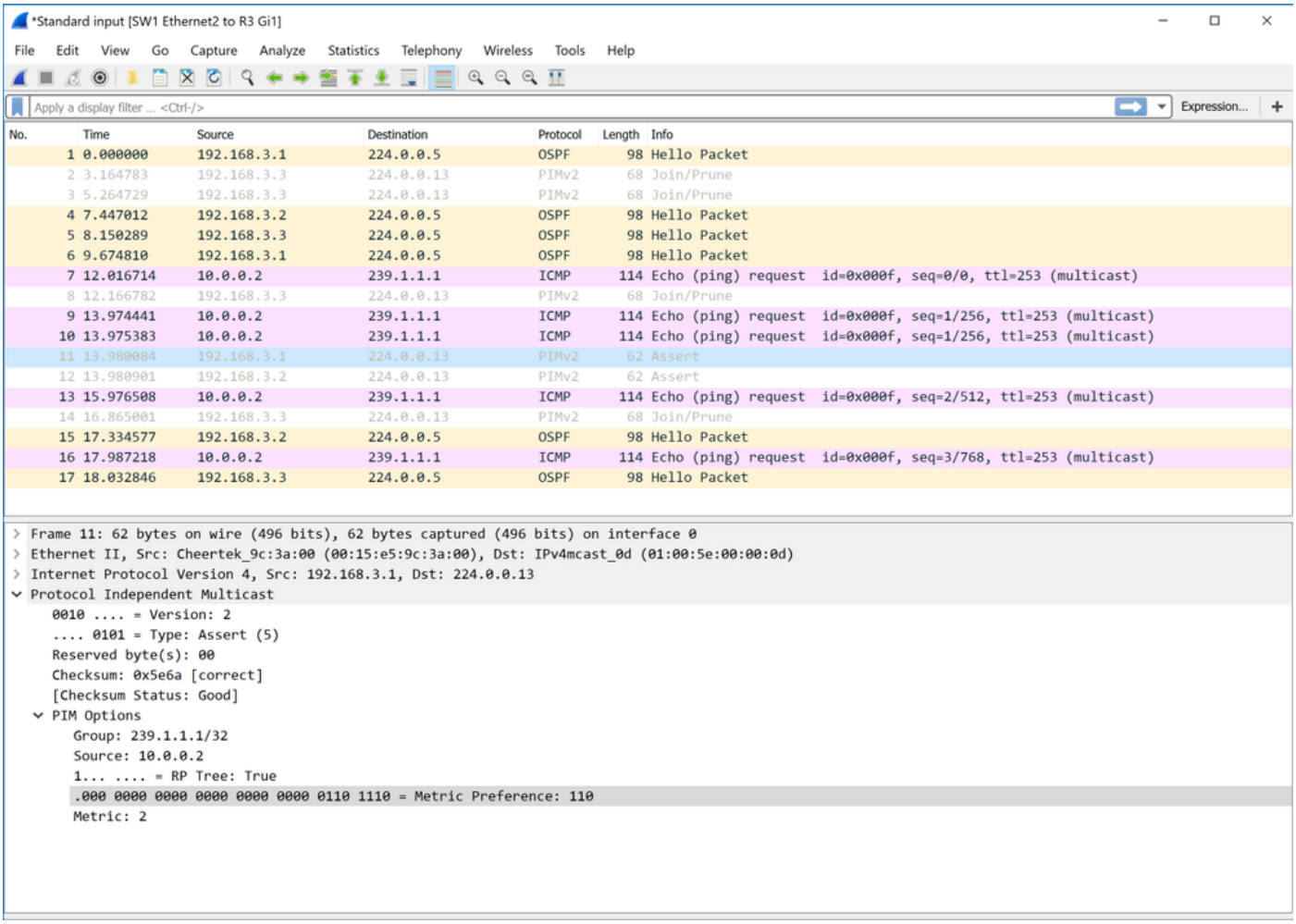
```
R3#show ip rpf 10.0.0.2
RPF information for ? (10.0.0.2)
  RPF interface: GigabitEthernet1
```

```
  RPF neighbor: ? (192.168.3.2)
  RPF route/mask: 10.0.0.0/24
  RPF type: unicast (ospf 1)
  Doing distance-preferred lookups across tables
  RPF topology: ipv4 multicast base, originated from ipv4 unicast base
```

Prima di attivare l'origine, diamo un'occhiata alla tabella mroute su R3, come si può vedere che c'è già (*,G) per il gruppo 239.1.1.1. Questo perché il ricevitore collegato a LHR ha già richiesto per il gruppo specificato.

```
R3#show ip mroute
IP Multicast Routing Table
Outgoing interface flags: H - Hardware switched, A - Assert winner, p - PIM Join
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 239.1.1.1), 00:00:57/00:02:32, RP 192.168.0.100, flags: S
  Incoming interface: GigabitEthernet1, RPF nbr 192.168.3.1
  Outgoing interface list:
    GigabitEthernet2, Forward/Sparse, 00:00:57/00:02:32

(*, 224.0.1.40), 00:11:24/00:02:41, RP 192.168.0.100, flags: SJCL
  Incoming interface: GigabitEthernet1, RPF nbr 192.168.3.1
  Outgoing interface list:
    GigabitEthernet2, Forward/Sparse, 00:02:02/00:02:41
```

A questo punto, attivare l'origine e acquisire i pacchetti sull'interfaccia R3 Gi1.



Come si può vedere in questa acquisizione, PIM afferma che i pacchetti sono già presenti.

Frame 11:

```
> Frame 11: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
> Ethernet II, Src: Cheertek_9c:3a:00 (00:15:e5:9c:3a:00), Dst: IPv4mcast_0d (01:00:5e:00:00:0d)
> Internet Protocol Version 4, Src: 192.168.3.1, Dst: 224.0.0.13
v Protocol Independent Multicast
    0010 .... = Version: 2
    .... 0101 = Type: Assert (5)
    Reserved byte(s): 00
    Checksum: 0x5e6a [correct]
    [Checksum Status: Good]
  v PIM Options
      Group: 239.1.1.1/32
      Source: 10.0.0.2
      1... .... = RP Tree: True
      .000 0000 0000 0000 0000 0000 0110 1110 = Metric Preference: 110
      Metric: 2
```

Frame 12

```
> Frame 12: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
> Ethernet II, Src: Cheertek_8b:3e:00 (00:15:e5:8b:3e:00), Dst: IPv4mcast_0d (01:00:5e:00:00:0d)
> Internet Protocol Version 4, Src: 192.168.3.2, Dst: 224.0.0.13
v Protocol Independent Multicast
    0010 .... = Version: 2
    .... 0101 = Type: Assert (5)
    Reserved byte(s): 00
    Checksum: 0xde6a [correct]
    [Checksum Status: Good]
  v PIM Options
      Group: 239.1.1.1/32
      Source: 10.0.0.2
      0... .... = RP Tree: False
      .000 0000 0000 0000 0000 0000 0110 1110 = Metric Preference: 110
      Metric: 2
```

Quando si esaminano questi pacchetti, si dovrebbe essere in grado di determinare chi è il vincitore assoluto. A questo punto si analizzerà la selezione del server di inoltro di asserzione PIM.

La preferenza della metrica è la distanza amministrativa (AD). Questo valore si riferisce alla distanza amministrativa del protocollo di routing che installa il percorso nella tabella di routing, utilizzata per cercare l'indirizzo IP di origine e la metrica è il costo del percorso.

Esistono anche altri attributi che vengono utilizzati per determinare chi è il vincitore dell'asserzione. Questi dettagli sono disponibili nella RFC 7761.

Riassunto della RFC 7761, sezione 4.6.3.

```
4.6.3.  Assert Metrics

   Assert metrics are defined as:

     struct assert_metric {
       rpt_bit_flag;
```

```
      metric_preference;
      route_metric;
      ip_address;
    };
  When comparing assert_metrics, the rpt_bit_flag, metric_preference,
  and route_metric fields are compared in order, where the first lower
  value wins.  If all fields are equal, the primary IP address of the
  router that sourced the Assert message is used as a tie-breaker, with
  the highest IP address winning.
```

Con l'utilizzo di questi campi definiti e la selezione del percorso, è possibile determinare chi vincerà in questo scenario. Se si osservano di nuovo i pacchetti di asserzione, si osserverà che la preferenza della metrica non viene confrontata poiché la decisione viene presa in base al primo criterio di selezione, ovvero rpt_bit_flag.

In questo scenario viene confrontato R1 e R2. Entrambi i router inviano messaggi di asserzione visualizzati in precedenza e, quando entrambi i dispositivi vedono i messaggi di asserzione l'uno dell'altro, possono confrontare le metriche tra loro per determinare chi è il vincitore.

Poiché R2 invia un messaggio di asserzione con l'albero RP: Se ha valore 0, è inferiore a quanto inviato da R1 con un albero RP: True, con il valore 1. Il bit dell'albero RP è impostato su 0 o 1.

Se impostato su 1, il bit dell'albero RP indica che l'utente si trova nella struttura condivisa; il bit RPT cancellato indica che il mittente dell'asserzione ha uno stato di inoltro (S,G) su un'interfaccia.

Poiché gli asserzioni (S,G) hanno priorità rispetto a (*,G), R2 dovrebbe essere il vincitore dell'asserzione. Come indicato nella dichiarazione precedente della RFC 7761, il valore più basso è quello più preferito.

Diamo un'occhiata sia a R1 che a R2 per vedere chi è il vincitore assoluto.

```
R2#show ip mroute
IP Multicast Routing Table
Outgoing interface flags: H - Hardware switched, A - Assert winner, p - PIM Join
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 239.1.1.1), 00:42:52/stopped, RP 192.168.0.100, flags: SP
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.4.1
  Outgoing interface list: Null

(10.0.0.2, 239.1.1.1), 00:42:52/00:01:40, flags: T
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.4.1
  Outgoing interface list:
    GigabitEthernet1, Forward/Sparse, 00:42:52/00:03:07, A

(*, 224.0.1.40), 00:43:23/00:02:25, RP 192.168.0.100, flags: SJPL
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.4.1
  Outgoing interface list: Null
```

In questo output, è possibile vedere che il (S,G) su R2 ha il flag A impostato su OIL che indica che è il vincitore dichiarato. In R1 non è presente un indicatore OIL su (S,G) e il flag P è impostato, il che significa che il particolare (S,G) è stato potato in questo caso: non è il vincitore assoluto.

> **Nota:** Quando un'asserzione è presente in un segmento condiviso, i vicini a valle inviano messaggi periodici Join(*,G) e Join(S,G) al router adiacente RPF appropriato, ovvero al

router adiacente RPF modificato dal processo di asserzione. Non sempre vengono inviate al router adiacente del RPF come indicato dal MRIB.

```
R1#show ip mroute
IP Multicast Routing Table
Outgoing interface flags: H - Hardware switched, A - Assert winner, p - PIM Join
 Timers: Uptime/Expires
 Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 239.1.1.1), 00:44:32/00:03:09, RP 192.168.0.100, flags: S
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.5.2
  Outgoing interface list:
    GigabitEthernet1, Forward/Sparse, 00:44:32/00:03:09, A

(10.0.0.2, 239.1.1.1), 00:44:19/00:03:09, flags: PR
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.5.2
  Outgoing interface list: Null

(*, 224.0.1.40), 00:44:50/00:02:53, RP 192.168.0.100, flags: SJCL
  Incoming interface: GigabitEthernet2, RPF nbr 192.168.5.2
  Outgoing interface list:
    GigabitEthernet1, Forward/Sparse, 00:43:56/00:02:53
```

Se in R1 e R2 il bit dell'albero RP è impostato su 1, è possibile prendere in considerazione il router con il bit AD più basso; se uguale, esaminare la metrica. Se il bit dell'albero RP è true su entrambi i router, la metrica viene confrontata con l'indirizzo IP RP. Se il bit dell'albero RP è 0, la metrica viene confrontata con l'origine del flusso multicast.

Se tutti questi valori sono uguali, il vincitore sarà il messaggio di asserzione di origine dell'indirizzo IP più alto.

Riepilogo

Nello scenario 1, non sono stati osservati pacchetti di asserzione, tuttavia, per RFC avrebbero dovuto essere attivati. Come accennato, questo avveniva perché R3 stava potando (*,G) prima che il control plane per (S,G) fosse costruito.

Nello scenario 2 vengono visualizzati pacchetti di asserzione. Quando il primo pacchetto è stato ricevuto su LHR, l'utente inviava un join/pruna (S,G) verso R3 per prelevare l'origine/il gruppo. R3 invia quindi un pacchetto di unione/eliminazione a R2 per la stessa origine/lo stesso gruppo. In questo modo, sia R1 che R2 conterrebbero degli OLI validi. Ora R3 elimina solo (S,G) con RP-bit impostato quando il flag T è inserito nello stato R3 (S,G). A tale scopo, è necessario ricevere un altro pacchetto del piano dati dal segmento condiviso. Poiché il control plane è già stato creato per (S,G), ciò determina la duplicazione sul segmento condiviso che attiva messaggi di asserzione.